Binary Decision Tree Using Genetic Algorithm for Recognizing Defect Patterns of Cold Mill Strip

Kyoung Min Kim[1,4], Joong Jo Park[2], Myung Hyun Song[3], In Cheol Kim[1], and Ching Y. Suen[1]

[1] Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, 1455 de Maisonneuve Blvd. West, Suite GM606, Montreal, Canada H3G 1M8
{kkm, kiminc, suen}@cenparmi.concordia.ca
[2] Department of Control and Instrumentation Engineering, Gyeongsang National University, 900, Gazwa-dong, Chinju, Gyeongnam, 660-701, Korea
[3] Department of Electric Control Engineering, Sunchon National University, 540-742, Korea
[4] Department of Electrical Engineering, Yosu National University, 550-749, Korea

Abstract. This paper presents a method to recognize the various defect patterns of a cold mill strip using a binary decision tree constructed by genetic algorithm(GA). In this paper, GA was used to select a subset of the suitable features at each node in the binary decision tree. The feature subset with maximum fitness is chosen and the patterns are divided into two classes using a linear decision function. In this way, the classifier using the binary decision tree can be constructed automatically, and the final recognizer is implemented by a neural network trained by standard patterns at each node. Experimental results are given to demonstrate the usefulness of the proposed scheme.

1 Introduction

To produce a cold mill strip of high quality, it is important to detect the defects on the surface of cold mill strip exactly in the manufacturing process. So, efficient methods for the recognition and extraction of defect patterns of cold mill strips have been studied [1, 8, 10].

The conventional method to recognize the defect patterns is to extract good features experimentally from the cold mill strip image acquired from a CCD camera and then recognize the patterns in a single step by inputting all the features to a neural network. But this method has two problems when the characteristics of the defect patterns are considered. Firstly, because the shapes of the defect patterns are complex and irregular, the recognition rate of defect patterns is sensitive to the kinds of selected features. And also despite the good separability of the features, they may interfere with each other when used together. Secondly, because there exist some similar classes of defect patterns, which can be classified into the same group, classifying all the patterns in only a single step results in a high classification error.

To overcome these problems, we propose a multi-stage classifier like a decision tree, which repeats decisions so as to classify patterns individually. The decision tree classifier makes fast and exact decisions by dividing the complex and global decisions

A.Y. Tawfik and S.D. Goodwin (Eds.): Canadian AI 2004, LNAI 3060, pp. 461–466, 2004.

into several simple and local decisions [5]. For an efficient and accurate classification, an optimal or near-optimal feature subset within the feature space needs to be selected at each decision node [2].

This paper introduces the binary decision tree and describes methods of both generating a linear decision function and selecting a feature subset using GA. Then, an automatic method of constructing the binary decision tree is described. And finally, the proposed classifier is applied to recognize the defect patterns of a cold mill strip.

2 Construction of Binary Decision Tree Using GA

In the case of a one-stage classifier, it cannot be guaranteed that all the features used in classification have the best separability, and the efficiency of the classifier is low because it compares a pattern with all of the other classes [6, 7]. To solve these problems, we have designed a classifier that decides the class of a given input pattern by repeating two or more decisions successively. This classifier is called a multi-stage classifier or a decision tree classifier.

If only the necessary features are used at each node of binary decision tree classifier, both the accuracy and the reliability of the classifier increase. So the problem is to select a valid feature subset from the entire feature set, i.e. the feature selection problem. To select the optimized feature subset, the separability of all the combinations of the features should be evaluated. However, when m components are selected among n components, the number of combinations expressed as nCm becomes a large value even if n and m are not large. The feature selection problem can be regarded as an optimization problem. So in this paper, feature selection is executed using GA, a global optimization technique.

GA has a higher probability of finding the global optimized solution than other classical optimization algorithms because it searches for the multiple global solutions simultaneously [1, 3, 4].

Fig. 1 shows the process of selecting the optimal feature subset by GA In this paper, a feature subset is evaluated by the classification error when classifying patterns with the linear decision function that is also generated by GA.

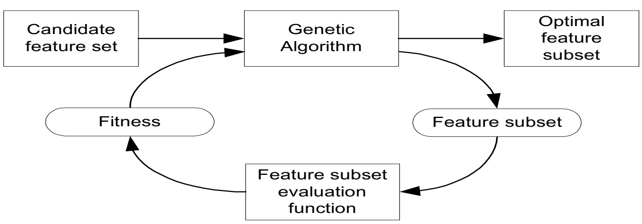


Fig. 1. Processing block diagram of feature selection

The following method is used to minimize the classification error using GA: Suppose that the given data set is $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ ($\mathbf{x}_k \in R^n$ is the number of features), and $l(j)$ and $r(j)$ are defined as the minimum and maximum values of the j-th feature.

$$l(j) = \min_i \ x_{ij} \tag{1}$$
$$r(j) = \max_i \ x_{ij}$$

In the case of 2-dimensional space, j can have the value of 1 or 2 and on the basis of $l(j)$ and $r(j)$, a rectangle can be constructed that can include all data. Inside the rectangle, two points can be selected arbitrarily, connected by a line. From the coefficients of the line function, a n -dimensional decision function can be obtained as follows.

$$d(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} = \mathbf{w}_0{}'\mathbf{x} + w_{n+1} \tag{2}$$

When matching this concept with a binary string of GA, $\boldsymbol{n}$ segments of a binary string indicate one point in the n-dimensional space. In the n-dimensional case, $\boldsymbol{n}$ points should be selected in such a way that a string is composed of n^2 segments. Supposing that the length of each segment is m-bit, then the total length of the binary string becomes $n^2 m$ -bits. The decision function described above is the linear decision function that minimizes the error that occurs when classifying standard patterns into two classes in a certain feature space. However, the real input patterns are not restricted to two classes of patterns. GA determines the decision function that minimizes classification error in a given feature space. Since the minimized error varies with the combination of features, the fitness function is constructed to give high fitness for a combination with a small classification error and low fitness for a combination with a large classification error.

Using the method described above, a certain feature subset minimizing classification error is chosen. And patterns are classified into two groups at each node with this feature subset. The binary decision tree is constructed by iterating this process until all the classes of patterns appear independently at each leaf node. Because the binary decision tree is constructed for multiple classes rather than just for two it is better to maintain uniform distribution for two separated groups at each node, which means it is better that two separated groups have similar numbers of classes without partiality.

To quantify this, a balance coefficient is defined using the mean and deviation of classes of a new group, as Eq. (3). If the number of patterns of the two separated groups is similar, the balance coefficients are smaller. In this case, because the depth of the binary tree becomes small, the matching time required for recognizing a pattern decreases.

$$balance = \sqrt{\frac{\sum_{j=1}^{h} (N_j - \frac{N}{h})^2}{(\frac{N}{h})^2}} \tag{3}$$

In Eq. (3), h is the number of nodes, N is the number of input patterns, and N_j is the number of the patterns included in the j-th node. In this paper, a binary tree is constructed, so h becomes $\boldsymbol{2}$. The fitness function that includes the balance coefficient is defined as.

$$fitness = \frac{1}{1 + w_e \cdot error + w_b \cdot balance} \tag{4}$$

In Eq. (4), *error* and *balance* are the classification error and the balance coefficient between groups, respectively, and w_e and w_b are the weights for weighting each parameter. And the result of the constructed tree can be varied by adjusting of the weights w_e and w_b. After the construction of the binary decision tree, by training BP neural network with the feature subset selected optimally at each node, the final binary tree structured recognizer is realized.

3 Classification of the Defects of Cold Mill Strip Using the Binary Tree Classifier

The defect patterns of cold mill strips can be classified into seven classes: Dull, Oil-drop, Slip, Dent, Scale, Dirt, and Scratch. After preprocessing for the acquired image, we extract six candidate features [8, 10].

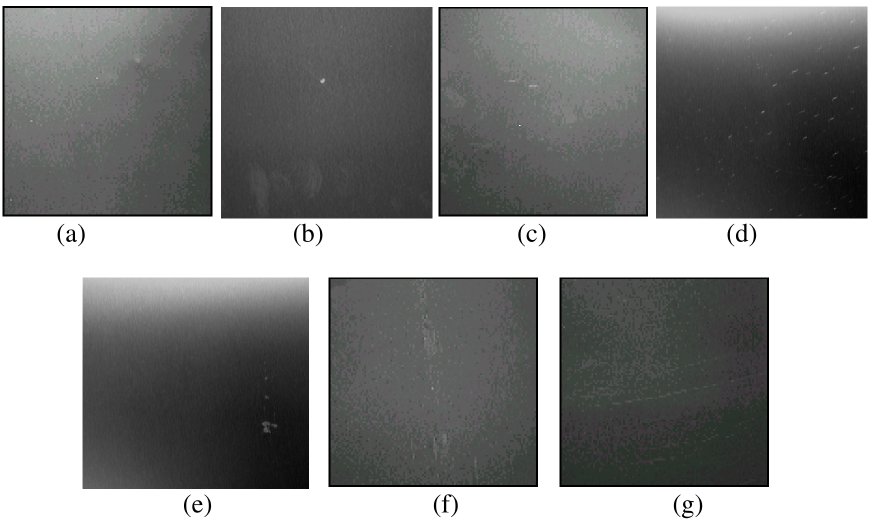
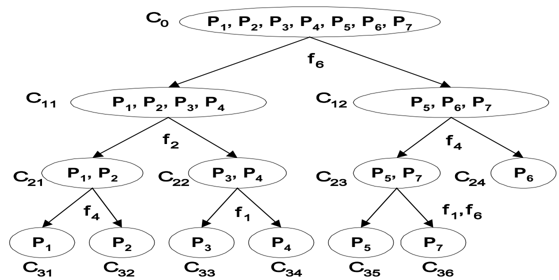


Fig. 2. Defect patterns of cold mill strip (a) dull (b) oil drop (c) slip (d) dent (e) dirt (f) scale (g) scratch

In this paper, geometrical features are selected as candidate features. They are area, area ratio, and compactness. They are not related to the size and direction of the patterns. And, the probabilistic concept of moment has been widely used in pattern recognition as a practical method to extract features of the shape. Among the moment-based features, the information useful for the inspection of the defects of cold mill strips are the length of the longest axis, the ratio of longest axis to shortest axis, and the spread of a pattern [8, 10]. The data used in constructing the binary tree recognizer are the feature vectors extracted from the seven types of standard defect patterns. In constructing the binary tree using GA, the weights in Eq. (4), w_e and w_b, are set to 1. Fig. 3 shows the binary decision tree constructed from standard patterns by GA. Here, P_i is a type of pattern, f_k is a feature, and C_{mn} represents a class at each node.



Symbol	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Pattern	dull	oil	slip	dent	scale	dirt	scratch

Symbol	f_1	f_2	f_3	f_4	f_5	f_6
Feature	def_area	area_ratio	compact-ness	axis_ratio	spread	long_axis

Fig. 3. Binary decision tree constructed with standard patterns

At each node constructed above, the final recognizer is made by training the BP neural network with the selected feature subset. The number of nodes in the input layer is set to the number of the selected features, and the number of nodes in the hidden layer is set to 10. By setting the number of nodes in the output layer to 2, the output layer represents the binary decision.

Table 1 shows the results of recognizing the defect patterns of a cold mill strip using the binary tree recognizer. In this table, the recognition rates of Dent and Slip are very low. However, the linear classification errors are zero at nodes C_0, C_{11}, and C_{22} when constructing the binary decision tree. This means that the standard patterns of Dent and Slip are classified linearly. Because the least number of features that fit to classify standard patterns are selected, if the number of standard patterns is small, the recognizer becomes sensitive to noise.

Table 1. Recognition rate of each defect pattern

Pattern	*No. of Recog. / No. of Patterns*	*Recognition rate (%)*
Dent	0/3	0.0
Dull	6/12	50.0
Oil drop	4/4	100.0
Slip	1/4	25.0
Dirt	2/2	100.0
Scale	19/22	86.4
Scratch	7/8	87.5
Total	39/55	70.9%

4 Conclusions

In this paper, we used a binary decision tree classifier to recognize the defect patterns of a cold mill strip. We have used the cold mill strip of POSCO (Pohang Steel Company), which consists of 55 defect patterns. At each node of the binary tree, GA was used for the selection of the best feature subset, and the linear decision function, also generated by GA, was used to calculate the fitness of a feature subset. There are two advantages of this method. One is that the construction of the binary decision tree and the selection of the best feature subset can be executed automatically for the given patterns. The other is that by designing the fitness function of GA properly, the decision tree can be obtained by considering the balance of the classes as well as the classification error.

Current performance is about 71% of recognition rate. Further studies should be made to design classifiers which have more generalization capabilities and feature extraction methods which are mutual helpful for the recognition of the defect pattern of a cold mill strip.

References

1. Jung S.W., Park G.T.: Design and Application of Binary Decision Tree using Genetic Algorithm. Journal of the KITE, 33(6) (1996) 1122-1130
2. Brill F., Brown D., Martin W.: Fast Genetic Selection of Features for Neural Network Classifiers. IEEE Trans. Neural Networks, 32 (1992) 324-328
3. Yao L.: Nonparametric Learning of Decision Regions via the Genetic Algorithm. IEEE Trans. Sys. Man Cybern., 26 (1996) 313-321
4. Mui J.K., Fu K.S.: Automated Classification of Nucleated Blood Cells using a Binary Tree Classifier. IEEE Trans. Pattern Analysis Machine Intelligence, 5 (1980) 429-443
5. Safavian S.R., Landgrebe D.: A Survey of Decision Tree Classifier Methodology. IEEE Trans. Sys. Man Cybern., 21 (1991) 660-674
6. Payne H.J., Meisel W.S.: An Algorithm for Constructing Optimal Binary Decision Trees. IEEE Trans. Computers, 26 (1997) 905-916
7. Swain P.H., Hauska H.: The Decision Tree Classifier: Design and Potential. IEEE Trans. Geosci. Elec., 15 (1997) 142-147
8. Kim K.M.: Development of Surface Defect Inspection Algorithms for Cold Mill Strip. Journal of Control, Automation and Systems Eng., 3 (1997) 179-186
9. Babu G.P., Murty M.N.: Clustering with Evolution Strategies. Pattern Recognition, 27 (1994) 321-329
10. Kim K.M.: Design of Hierarchical Classifier for Classifying Defects of Cold Mill Strip using Neural Network. Journal of Control, Automation and Systems Eng., 4 (1998) 499-505